



US008538984B1

(12) **United States Patent**  
**Mahabal et al.**

(10) **Patent No.:** **US 8,538,984 B1**  
(45) **Date of Patent:** **Sep. 17, 2013**

(54) **SYNONYM IDENTIFICATION BASED ON CO-OCCURRING TERMS**

(75) Inventors: **Abhijit A. Mahabal**, Jersey City, NJ (US); **Takahiro Nakajima**, Chiba (JP); **Zachary A. Garrett**, Tokyo (JP); **Kenji Inoue**, Tokyo (JP)

(73) Assignee: **Google Inc.**, Mountain View, CA (US)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **13/438,125**

(22) Filed: **Apr. 3, 2012**

(51) **Int. Cl.**  
**G06F 17/30** (2006.01)

(52) **U.S. Cl.**  
USPC ..... **707/769**

(58) **Field of Classification Search**  
USPC ..... 707/2, 3, 769; 709/219  
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

7,636,714 B1 12/2009 Lamping et al.  
2004/0186831 A1 \* 9/2004 Hiratsuka et al. .... 707/3  
2012/0166473 A1 \* 6/2012 Ahmadi et al. .... 707/769

\* cited by examiner

*Primary Examiner* — Etienne LeRoux

(74) *Attorney, Agent, or Firm* — Fish & Richardson P.C.

(57) **ABSTRACT**

Methods, systems, and apparatus, including computer programs encoded on a computer storage medium, for identifying a particular query term of an original search query, identifying a candidate synonym for the particular query term in context with an other non-adjacent query term of the original search query that is not adjacent to the particular query term in the original search query, accessing stored data that specifies, for a pair of terms that includes the particular query term and the candidate synonym of the particular query term, a respective confidence value for the other non-adjacent query term, determining that, in the stored data, the confidence value for the other non-adjacent query term satisfies a threshold, and determining to revise the original search query to include the candidate synonym of the particular query term, based on determining that the confidence value the other non-adjacent query term satisfies the threshold.

**26 Claims, 6 Drawing Sheets**

